

# **Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review**

Year: 2021 | Citations: 456 | Authors: A. Antoniadi, Yuhan Du, Yasmine Guendouz, Lan Wei, Claudia Mazo

---

## **Abstract**

Machine Learning and Artificial Intelligence (AI) more broadly have great immediate and future potential for transforming almost all aspects of medicine. However, in many applications, even outside medicine, a lack of transparency in AI applications has become increasingly problematic. This is particularly pronounced where users need to interpret the output of AI systems. Explainable AI (XAI) provides a rationale that allows users to understand why a system has produced a given output. The output can then be interpreted within a given context. One area that is in great need of XAI is that of Clinical Decision Support Systems (CDSSs). These systems support medical practitioners in their clinic decision-making and in the absence of explainability may lead to issues of under or over-reliance. Providing explanations for how recommendations are arrived at will allow practitioners to make more nuanced, and in some cases, life-saving decisions. The need for XAI in CDSS, and the medical field in general, is amplified by the need for ethical and fair decision-making and the fact that AI trained with historical data can be a reinforcement agent of historical actions and biases that should be uncovered. We performed a systematic literature review of work to-date in the application of XAI in CDSS. Tabular data processing XAI-enabled systems are the most common, while XAI-enabled CDSS for text analysis are the least common in literature. There is more interest in developers for the provision of local explanations, while there was almost a balance between post-hoc and ante-hoc explanations, as well as between model-specific and model-agnostic techniques. Studies reported benefits of the use of XAI such as the fact that it could enhance decision confidence for clinicians, or generate the hypothesis about causality, which ultimately leads to increased trustworthiness and acceptability of the system and potential for its incorporation in the clinical workflow. However, we found an overall distinct lack of application of XAI in the context of CDSS and, in particular, a lack of user studies exploring the needs of clinicians. We propose some guidelines for the implementation of XAI in CDSS and explore some opportunities, challenges, and future research needs.